Belief propagation in imprecise Markov trees

Filip Hermans, Gert de Cooman & Erik Quaeghebeur

SYSTeMS Research Group, EESA Department, Faculty of Engineering, Ghent University
Technologiepark-Zwijnaarde 914, 9052 Zwijnaarde

{Filip.Hermans,Gert.deCooman,Erik.Quaeghebeur}@UGent.be

Basic notions and notation

We consider a rooted and directed discrete tree with finite width and depth:

- we call T the set of its *nodes* s
- we denote the set of *parents* of node s by $P(s)$; if $P(s)$ is non-empty then $P(s)=\{m(s)\}$, with $m(s)$ the *mother* of s
- we denote the set of *children* of s by $C(s)$, and the set of its *siblings* by $S(s)$
- we write $s \sqsubseteq t$ if node s *precedes* node t
- $D(s) := \{t \in T : s \sqsubset t\}$ denotes the set of *descendants* of s, and $A(s) := \{t \in T : t \sqsubset s\}$ its set of *ancestors*; $s \sqsubset t$ means that $s \sqsubseteq t$ and $s \neq t$
- we also use $\downarrow s := D(s) \cup \{s\}$ and $\uparrow s := A(s) \cup \{s\}$

With each node s of the tree, there is associated a variable X_s assuming values in a finite non-empty set $\mathscr{X}_s$.

Local uncertainty models

We now add a *local uncertainty model* to each of the nodes:

- a conditional lower expectation $\underline{Q}_s(\cdot|X_{m(s)})$ on $\mathbb{R}^{\mathscr{X}_s}$: for each possible mother-value $z_{m(s)}$, we have a lower expectation $\underline{Q}_s(\cdot|z_{m(s)})$.
- an unconditional lower expectation $\underline{Q}_\square$ on $\mathbb{R}^{\mathscr{X}_\square}$.

We use the common generic notation $\underline{Q}_s(\cdot|X_{P(s)})$ for all these local belief models.

Lower expectations Instead of specifying a single (precise) probability model, an expert expresses his beliefs by giving bounds on this belief model. For example: "the probability of A is higher than that of B". This leads to a convex closed set of precise probability models, a so-called *credal set* $\mathscr{P}$.

Specifying a convex set $\mathscr{P}$ of probability mass functions p on finite set $\mathscr{X}$ is equivalent to specifying *lower* and *upper expectations*, defined for any $g \in \mathbb{R}^{\mathscr{X}}$ by

$$\underline{E}(g) := \min\left\{\sum_{x\in\mathscr{X}} g(x)p(x) \colon p \in \mathscr{P}\right\},$$

$$\overline{E}(g) := \max\left\{\sum_{x\in\mathscr{X}} g(x)p(x) \colon p \in \mathscr{P}\right\}.$$

Observe the *conjugacy relationship*:

$$\overline{E}(g) = -\underline{E}(-g).$$

The real functional $\underline{E}$ is bounded, non-negatively homogeneous and super-additive. There is a one-to-one relationship between credal sets and lower expectation functionals.

Interpretation of the graphical model

Epistemic irrelevance Y *is irrelevant to* X whenever the belief model (lower expectation $\underline{E}$) about X does not change when we learn something about Y:

$$(\forall g \in \mathbb{R}^{\mathscr{X}})(\forall y \in \mathscr{Y})\underline{E}(g) = \underline{E}(g|y).$$

It is not symmetrical and does not imply d-separation in trees.

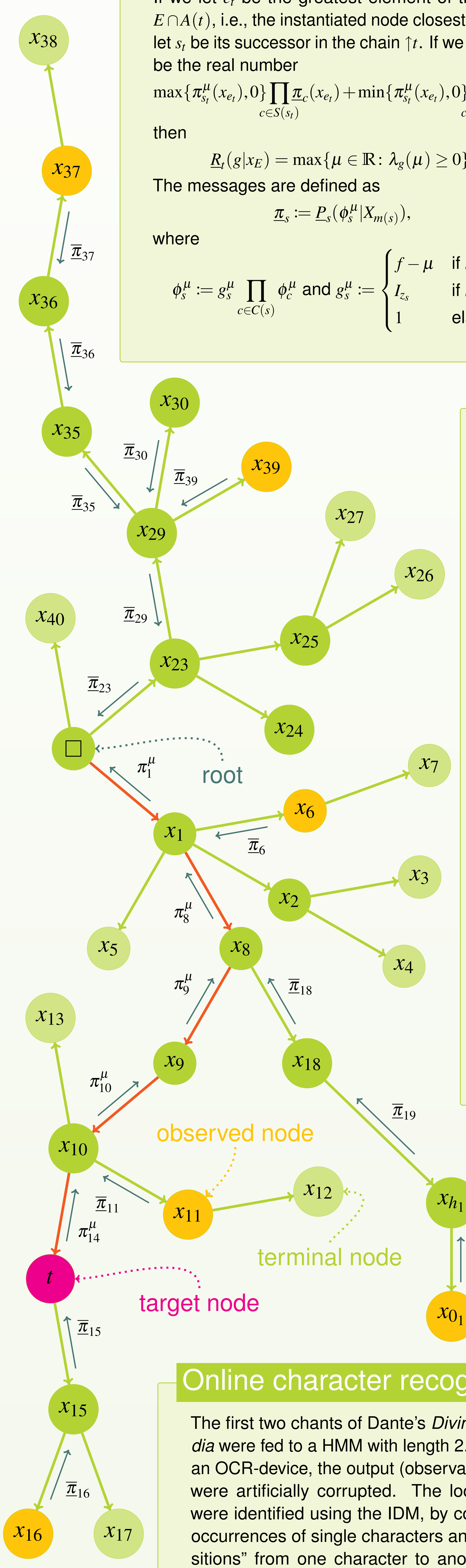
Interpretation of the graphical structure Consider any node s, its mother $m(s)$ and the set $\bar{s} := T \setminus [D(s) \cup \{m(s)\}]$ of the non-parent non-descendants of s. Then *conditional on the mother variable $X_{m(s)}$, the non-parent non-descendant variables $X_{\bar{s}}$ are assumed to be epistemically irrelevant to the variables $X_{\downarrow s}$ associated with s and its descendants.*

This means that for all $s \in T$, for all $S \subseteq \bar{s}$ and for all $z_{S\cup P(s)} \in \mathscr{X}_{S\cup P(s)}$:

$$\underline{P}_s(\cdot|z_P) = \underline{P}_s(\cdot|z_{S\cup P}).$$

This makes the tree an *imprecise Markov tree* (IMT).

Recursive construction of the joint Using the interpretation of the graphical structure, and the local belief models $\underline{Q}_s(\cdot|X_{P(s)})$, we can construct the most conservative joint lower expectation $\underline{P}$ for all variables in the tree in a recursive fashion, from leaves to root.



Belief updating

We treat the imprecise Markov tree as an expert system, i.e. we are interested in making inferences about the value of the variable X_t in some *target node* t, when we know the values x_E of the variables X_E in a set $E \subseteq T \setminus \{t\}$ of *evidence nodes*. Assuming that $\overline{P}(\{x_E\}) > 0$, we can do this by conditioning the joint $\underline{P}$ on the available evidence '$X_E = x_E$'.

$$\underline{R}_t(g|x_E) = \max\{\mu \in \mathbb{R} \colon \underline{P}(I_{\{x_E\}}[g-\mu]) \geq 0\}.$$

If we let e_t be the greatest element of the chain $E \cap A(t)$, i.e., the instantiated node closest to t, and let s_t be its successor in the chain $\uparrow t$. If we let $\lambda_g(\mu)$ be the real number

$$\max\{\pi_{s_t}^\mu(x_{e_t}),0\}\prod_{c\in S(s_t)}\underline{\pi}_c(x_{e_t}) + \min\{\pi_{s_t}^\mu(x_{e_t}),0\}\prod_{c\in S(s_t)}\overline{\pi}_c(x_{e_t}),$$

then

$$\underline{R}_t(g|x_E) = \max\{\mu \in \mathbb{R} \colon \lambda_g(\mu) \geq 0\}.$$

The messages are defined as

$$\underline{\pi}_s := \underline{P}_s(\phi_s^\mu|X_{m(s)}),$$

where

$$\phi_s^\mu := g_s^\mu \prod_{c\in C(s)} \phi_c^\mu \text{ and } g_s^\mu := \begin{cases} f-\mu & \text{if } s = \square, \\ I_{z_s} & \text{if } s \in E, \\ 1 & \text{else.} \end{cases}$$

The messages $\underline{\pi}_{s_t}, \overline{\pi}_{s_t}$ and $\pi_{s_t}^\mu$ can be computed in a recursive way:

$$\overline{\underline{\pi}}_s = \begin{cases} \overline{\underline{Q}}_s\left(\{x_s\}|X_{m(s)}\right)\prod_{c\in C(s)}\overline{\underline{\pi}}_c(x_s) & \text{if } s \in E,\ s \not\sqsubseteq t \\ \overline{\underline{Q}}_s\left(\prod_{c\in C(s)}\overline{\underline{\pi}}_c|X_{m(s)}\right) & \text{if } s \notin E,\ s \not\sqsubseteq t. \end{cases}$$

The messages $\underline{\pi}_s$ and $\overline{\pi}_s$ can be seen as tuples of real numbers, with as many components as there are elements in $\mathscr{X}_{m(s)}$: one for each of the possible values of $X_{m(s)}$.
Next, we turn to nodes $s \sqsubseteq t$. Define the messages π_s^μ by

$$\pi_s^\mu := \underline{Q}_s(\psi_s^\mu|X_{P(s)}),$$

where the gambles ψ_s^μ on $\mathscr{X}_s$ are given by the recursion relations:

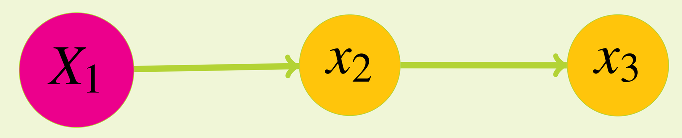
$$\psi_t^\mu := \max\{g-\mu,0\}\prod_{c\in C(t)}\underline{\pi}_c + \min\{g-\mu,0\}\prod_{c\in C(t)}\overline{\pi}_c,$$

and for each $\square \neq s \sqsubseteq t$, so $m(s)$ exists,

$$\psi_{m(s)}^\mu := g_{m(s)}^\mu\left[\max\{\pi_s^\mu,0\}\prod_{c\in S(s)}\underline{\pi}_c + \min\{\pi_s^\mu,0\}\prod_{c\in S(s)}\overline{\pi}_c\right].$$

A simple example involving dilation

Consider the following imprecise Markov chain:



Suppose that $\mathscr{X}_1 = \{a,b\}$, $\underline{Q}_1$ is a linear model Q_1 with mass function q and that $\underline{Q}_2(\cdot|X_1)$ is a linear model $Q_2(\cdot|X_1)$ with conditional mass function $q(\cdot|X_1)$. We make no restrictions on the local model $\underline{Q}_3(\cdot|X_2)$.
We find after applying the belief updating that

$$\underline{r} := \underline{R}_1(\{a\}|x_{\{2,3\}}) = \frac{q(a)q(x_2|a)\underline{q}}{q(a)q(x_2|a)\underline{q} + q(b)q(x_2|b)\overline{q}}$$

$$\overline{r} := \overline{R}_1(\{a\}|x_{\{2,3\}}) = \frac{q(a)q(x_2|a)\overline{q}}{q(a)q(x_2|a)\overline{q} + q(b)q(x_2|b)\underline{q}}.$$

When $\underline{q} = \overline{q}$, which happens for instance if the local model for X_3 is precise, then we see that, with obvious notations,

$$\overline{r} = \underline{r} = \frac{q(a)q(x_2|a)}{q(a)q(x_2|a) + q(b)q(x_2|b)} =: p(a|x_2)$$

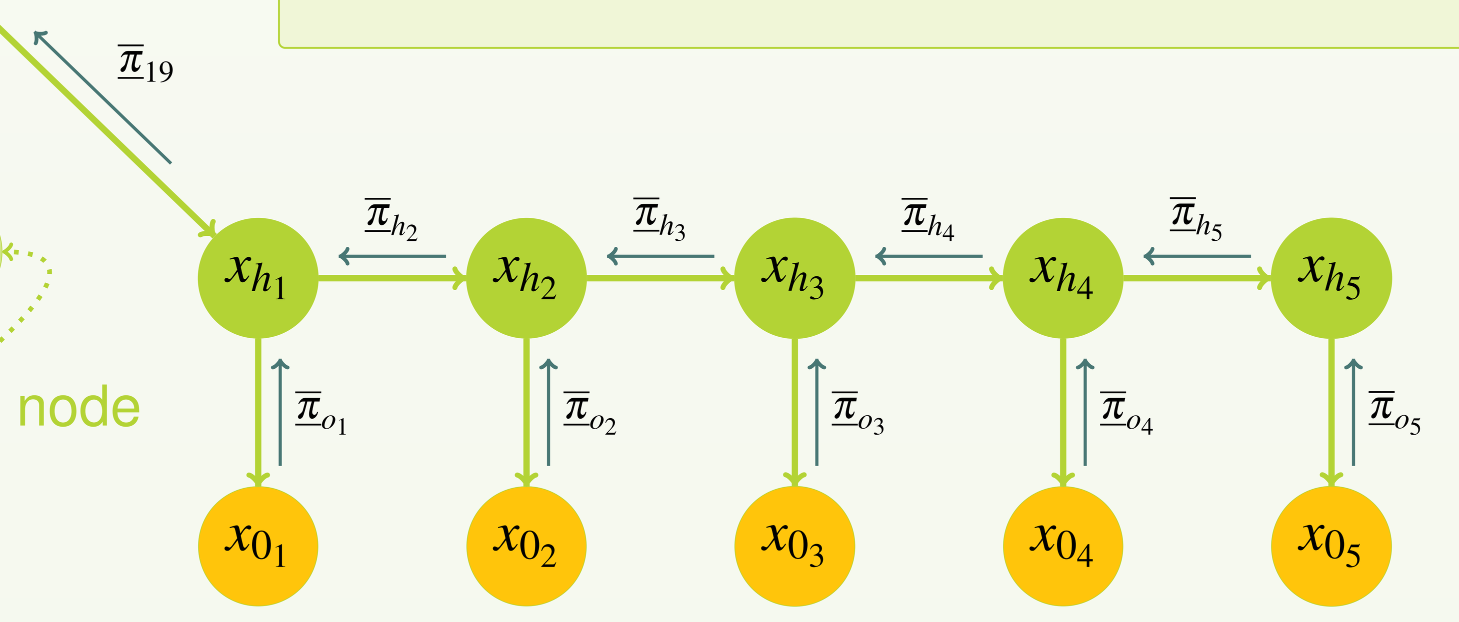
and therefore X_2 indeed separates X_3 from X_1. But in general, letting $\alpha := q(a)q(x_2|a)$ and $\beta := q(b)q(x_2|b)$, we get

$$\overline{r} - \underline{r} = \frac{\alpha\beta(\overline{q}^2 - \underline{q}^2)}{(\alpha^2+\beta^2)\underline{q}\overline{q} + \alpha\beta(\underline{q}^2 + \overline{q}^2)}$$

$$\overline{r} - p(a|x_2) = \frac{\alpha\beta}{\alpha+\beta}\frac{\overline{q}-\underline{q}}{\alpha\overline{q}+\beta\underline{q}}$$

$$p(a|x_2) - \underline{r} = \frac{\alpha\beta}{\alpha+\beta}\frac{\overline{q}-\underline{q}}{\alpha\underline{q}+\beta\overline{q}}.$$

As soon as $\overline{q} > \underline{q}$, X_2 no longer separates X_3 from X_1, and we witness *dilation* because of the additional observation of X_3!



Online character recognition by imprecise HMMs

The first two chants of Dante's *Divina Commedia* were fed to a HMM with length 2. Mimicking an OCR-device, the output (observation nodes) were artificially corrupted. The local models were identified using the IDM, by counting the occurrences of single characters and the "transitions" from one character to another in the original text.

Accuracy	93.96%	(7275/7743)
Accuracy (if imprecise indeterminate)	64.97%	(243/374)
Determinacy	95.17%	(7369/7743)
Set-accuracy	93.58%	(350/374)
Single accuracy	95.43%	(7032/7369)
Indeterminate output size	2.97	over 21